Computational Modelling of Reactivity in Zn-Containing Enzymes

Jon I. Mujika, Adrian J. Mulholland and Jeremy N. Harvey*
School of Chemistry and Centre for Computational Chemistry, University of Bristol, Cantock's Close, Bristol BS8 1TS, UK

ABSTRACT

Computation is increasingly well placed to offer insight into a range of structural and spectroscopic properties of Zn-containing enzymes, as well as into their reactivity. The latter aspect is the focus of this chapter. Molecular dynamics simulations provide insight into substrate binding and active site relaxation around the metal and substrate. A significant challenge in such studies is the representation of the zinc centre. Quantum mechanical methods, either on their own using small active site models, or for whole enzymes using hybrid techniques, provide insight into reaction paths and hence into transition states and intermediates. Examples of these applications are reviewed for carbonic anhydrase, carboxypeptidase A, thermolysin, type B β-lactamase, and farnesyltransferase.

1. INTRODUCTION

Many enzymes incorporate a metal ion cofactor, in order to take advantage of the unique chemical reactivity that such centres can confer.[1] After iron, zinc is the second most abundant transition metal in biological systems.[2] The ubiquity of Zn is due to a number of factors related to the toxicity and bioavailability of this ion, but also to some of its chemical characteristics that make it very well suited to serve as a reactive centre in enzymatic processes.[3] In particular, it has a flexible coordination sphere which can therefore change along the reaction path, and, as a small ion, it has a fairly large charge to radius ratio that make it a fairly strong Lewis acid. It is to be noted that unlike iron, the closed-shell configuration of the +2 oxidation state of Zn leads to a strong preference for this oxidation state, so that this ion does not show any redox activity.

There are several different types of Zn-containing enzymes with different biological functions (*see* IA259-).[4, 5] Based on the role of the zinc ion, the enzymes can be classified into three main types,[6] i.e. with structural, catalytic, and cocatalytic cofactors. In structural Zn enzymes, the metal is involved only in stabilizing the enzyme scaffold and has no direct catalytic role. Structural zinc sites involve four protein groups serving as metal ligands, with Cys side-chain thiolates the most common group of this type. In the other two classes of zinc enzymes, the ion plays an active catalytic role in the enzymatic reaction. Catalytic zinc sites involve a single ion, and here too, tetracoordinate zinc is the most common form, with a water molecule serving as one of the ligands in almost all cases, together with three protein groups which can include nitrogen, oxygen and sulfur donor atoms. The imidazole side chain of His residues is the most common ancillary ligand. In cocatalytic sites, there are two or three metal ions in close proximity, and two of the metals are bridged by a side-chain moiety of a single amino acid residue, such as an Asp or Glu carboxylate, a His deprotonated imidazole, or in some cases a deprotonated water molecule or hydroxide ion.

Understanding enzyme mechanisms requires knowledge of how transition states and intermediates are formed and stabilized within enzymes.[7] In order to understand how this stabilisation occurs, a detailed description of the system at the atomic level is required. Experimentally, many techniques such as X-ray crystal structures, NMR studies (*see* IA319-), linear free energy relationships (LFER), kinetic isotope effects or site-directed mutagenesis provide invaluable information about the reaction. However, each of these methods has limitations that prevent them from giving a full understanding of the nature of catalysis in a given enzymatic system. In this context, the contribution of computational chemistry is crucial as it can provide valuable complementary information. Calculations can be used to explore extended regions of the potential energy surfaces, not just those relating to the kinetically observable steps. They also lead to predicted structures (or dynamical ensembles) and properties for reaction intermediates, substrates and, most importantly, transition states, that may be very elusive to experimental characterization.

Theoretical and modelling approaches have been applied extensively in the study of zinc-containing enzymes using very diverse theoretical methods and addressing equally diverse aspects of zinc-enzyme chemistry. In this work, we will focus on the studies that address the reactivity of zinc-containing enzymes, where the zinc ion plays an active catalytic role in the reaction. We will therefore not cover in great detail issues such as the preferred coordination mode of Zn,[8] or the prediction of metal- and ligand-binding affinities. The review will start with a brief description of the computational techniques used, then move on to summarize some of the key computational work in the area, organized according to the nature of the target enzyme.

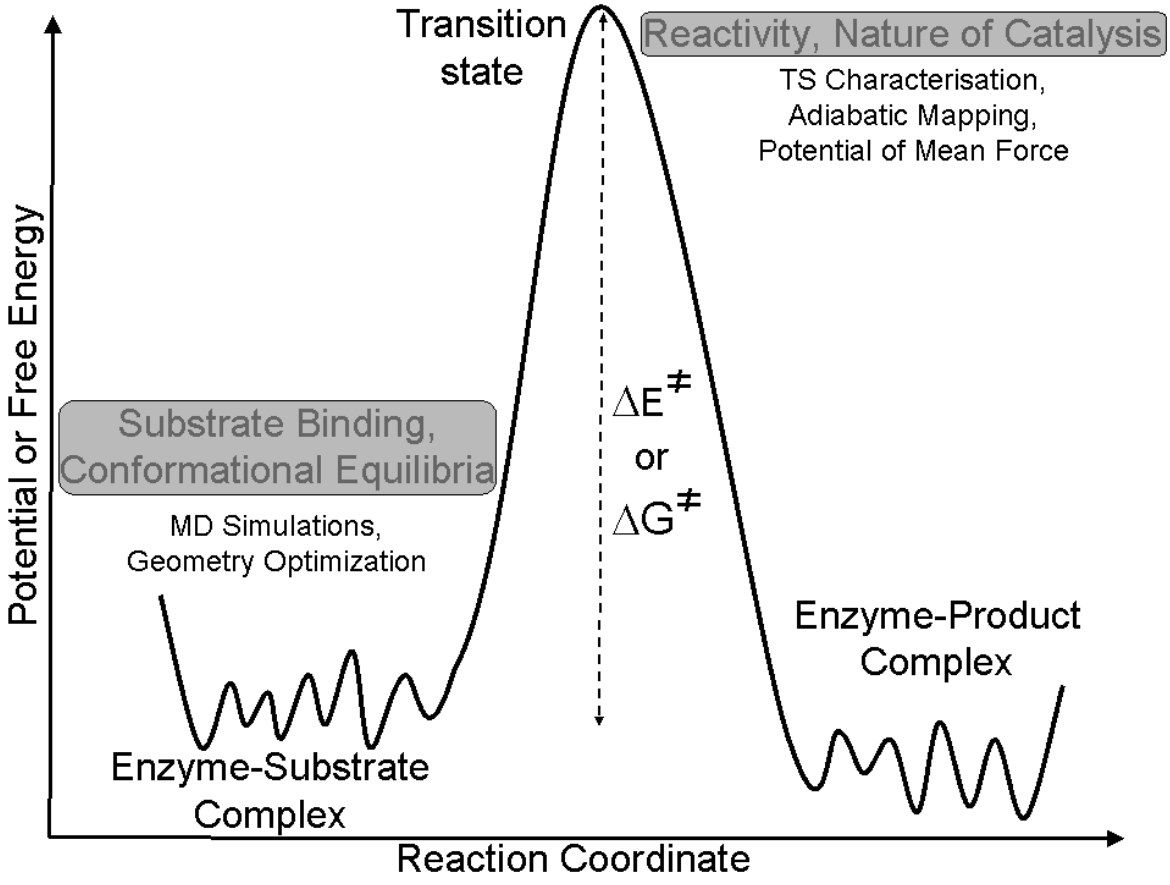

Figure 1 Generic potential or free energy surface for an enzyme-catalyzed reaction.

2. COMPUTATIONAL TECHNIQUES

The present section is designed to put in context the computational studies of zinc metalloenzymes discussed below. It is very brief, as more detailed explanations of the methods may be found in other chapters of the Encyclopedia or elsewhere in the literature (e.g., *see* IA609-).[9-11]

The aim of most computational studies of reactivity is an accurate characterization of either the potential energy surface for reaction, or the corresponding free energy surface (or both, see Figure 1). For enzymatic systems including thousands of atoms, the complete characterization of these energy surfaces is of course impossible. Depending on the property to be modelled, and the enzyme target, different approaches to constructing a tractable model are available. Among them, the most commonly used methods to study enzymatic reactivity can be divided into two types of calculations (Figure 1): molecular dynamics simulation based on molecular mechanics, and quantum mechanics based studies of reactivity (e.g., *see* IA652-, IA606-).

2.1 Molecular Dynamics simulations:

Enzymes are large molecules that may adopt a huge number of different conformations. Molecular dynamics (MD) simulations provide a means to sample the time-dependent and time-averaged behaviour associated with these many conformations. Within the limits associated with the physical timescales involved (for time-dependent phenomena) and with the statistical mechanical challenges, this type of calculation provides insight into a range of properties, with structure and dynamical behaviour of typical interest.

MD simulations require a force field, with parameters for terms such as bond stretching and bending, torsions, and non-bonded van der Waals and electrostatic interactions. Various MM force fields and programs have been developed for simulations of proteins, nucleic acids and lipids or other biological molecules. Compared to the quantum chemical calculations also discussed in this chapter, MD calculations are fast and allow reasonably broad sampling of even large systems (of thousands or tens of thousands of atoms). However, it should be borne in mind that there are some limitations inherent to the method. First, despite the existence of standard well-validated force fields, specific parameterisation and validation is often required if the system to be studied includes unusual substrates or inhibitors. Moreover this method does not straightforwardly allow the study of chemical reactions, since a single force-field cannot readily describe both the reactants and products when formation and breaking of chemical bonds occurs.

Treatment of Zn in MM MD simulations:

Much attention has been paid to the description of Zn ions in force fields. Although there is one simplifying aspect, in that the oxidation state of Zn is fixed, a wide variety of coordination modes are possible; in solution, the ion adopts a six-coordinate octahedral structure, while in a protein environment it is more often tetracoordinate (tetrahedral) or pentacoordinate (trigonal bipyramid or square pyramid).[8] The geometry around a Zn may change during a catalytic reaction, and this flexible coordination makes it difficult to obtain a good description of the Zn

coordination sphere using an empirical force field, and this should be remembered when assessing the results of the MD simulation studies described below.

Several types of MM model have been used to model Zn and its ligand sphere in enzymes. These models fall into three categories: bonded, non-bonded and dummy cationic models. In "bonded models", the structure of the metal binding site is preserved by introducing explicit bond and angle terms into the potential energy function to account for interactions between the metal and its ligands.[12] This technique enforces the chosen coordination mode, which can be an advantage in some respects, but it also prevents exchange or permutation of ligands, or changes in coordination during the simulation. Therefore, great care is needed when using this method.

"Non-bonded" models, in contrast, include only van der Waals and electrostatic interactions between the metal and other atoms, with no covalent terms at all. A model of this type was tested by running several MD simulations of two different zinc metalloenzymes, carboxypeptidase A and carbonic anhydrase.[13] For up to 101 ps, the coordination mode of the metal was observed to remain close to that found in the initial crystal structure. In later simulations over longer timescales, however, this model failed to preserve the tetrahedral configuration found in the X-ray crystal structure.[14] In enhanced non-bonded models,[14, 15] with so-called "cationic dummy atoms", some degree of directionality in bonding is introduced by treating the ion as a central charge surrounded by a set of virtual "dummy" charges.

A key factor for MM treatment of zinc is the charge used on the Zn ion. In most of the classical force fields, Zn bears a +2 charge, that is, charge transfer from the ligands is not considered. In more sophisticated approaches, however, such effects can be treated. For example, within a non-bonded model, it has been shown that good results can be obtained simply by assigning refined atomic charges based on density functional theory (DFT) calculations of a model of the metal centre.[16] In spite of its simplicity, this method led to a much improved stability of the coordination geometry in long MD simulations of a variety of metalloproteins. Improved results can also be obtained by e.g. including terms describing polarization and charge transfer between the metal and the surrounding ligands.[17, 18]

2.2 Reactivity studies using QM methods:

Quantum mechanical (QM) electronic structure calculations are often used in the study of chemical reactions, with popular methods based on ab initio wave function theory, DFT, and semi-empirical methods. In these approaches, electrons are treated explicitly, allowing for breaking and forming of chemical bonds. These calculations can be much more accurate than MM methods but are also more time-consuming, which restricts the size of system that can be studied. Different QM methods involve different approximations, yield different accuracy, and are of different computational expense. For Zn enzymes, both rapid but less accurate semi-empirical and more expensive yet more accurate DFT methods have been frequently used.

Due to computational expense, purely QM methods can only be used for relatively small 'cluster' models of the active site of the system. In such calculations, one can optimise the geometry of stationary points such as reactant complexes, intermediates, product complexes, and also transition states. Using transition state theory, the relative energy of the TS and the reactant complex can be used to compare predicted and experimental reactivity. However, such calculations usually only

characterize a single conformation for each stationary point, while the real potential energy surface of the protein system is characterized by extreme conformational variability (as shown schematically in Fig. 1). Moreover, as only a small number of atoms around the active site is considered, the influence of the environment is not taken into account. The problems can be alleviated to some extent by restricting some to their position in the crystal structure, or by using a larger model. These calculations can be very useful for discriminating between alternative possible mechanisms, especially if the difference in the calculated barrier heights is large.

An increasingly popular alternative to QM calculations on small cluster models is hybrid quantum mechanical/molecular mechanical (QM/MM) methods (*see* IA606-).[11, 19] These combine the accuracy of QM methods, used for the active site, with the speed and simplicity of MM methods, used for the rest of the system (protein and solvent). Reaction profiles can be derived in which the geometry of stationary points, including transition states, are fully optimised, with a computational expense similar to that of a QM calculation on the QM region only.

As with all molecular modelling, QM/MM techniques have some important limitations. First of all, large systems can adopt many different conformations, and an accurate characterization and sampling of the energy surface, especially transition states, may be difficult to achieve. Moreover, as for any QM calculation, there is a trade-off between the accuracy of the method used to treat the QM part and its computational expense. Where extensive sampling of the potential energy surface is required, therefore, many QM/MM studies use relatively simple QM levels of theory such as semi-empirical methods.

Saying that a calculation uses a QM/MM method merely specifies the technique used to calculate the potential energy for a given geometry. To study enzyme reactivity, one also needs to characterize the reactant complex, possible intermediates, and the relevant transition states. Two main techniques are commonly used with QM/MM methods for this purpose: adiabatic mapping, and potential of mean force calculations. In adiabatic mapping,[20] a reaction coordinate is chosen, that is believed to vary smoothly between the reactant, the TS, and the product for the step being considered. Geometry optimization is carried out for several constrained values of this coordinate, such that the reactant, product, and TS regions are covered. The structure with the highest energy can be considered to be the TS. This technique is fairly rapid, requiring only a handful of geometry optimizations, which in turn, with an efficient QM/MM implementation, have a similar computational expense to that of a QM geometry optimization. However, the results obtained with this approach can be highly dependent on the initial geometry chosen. Some initial geometries may not yield a meaningful reaction path, and some choices of reaction coordinate may also lead to poor results. Moreover, the characterization of the transition state is not exact. Other methods for obtaining minimum energy pathways can also be used.

In potential of mean force (PMF) or umbrella sampling calculations,[21] several MD simulations are performed, again using a predefined reaction coordinate. In successive simulations, the coordinate is constrained to lie close to certain reaction coordinate values, so that the set of simulations covers not only the reactant and product wells, but also the higher-energy regions along the reaction path and near the TS. A statistical treatment[22] is then used to obtain the free energy profile. PMF techniques are in principle much more accurate than adiabatic mapping calculations, because they allow for sampling of a range of conformations. However, to carry out the underlying MD simulations, it is necessary to carry out a very large number of calculations of the QM/MM energy and its gradient, and this makes these calculations

very time consuming. We note that there are many different free energy methods of this type that broadly fit the description given above, but the detailed features of each of them will not be discussed here.

2.3 REACTIVITY: ZN ENZYME CATALYSIS

In this section we will provide an overview of computational studies of zinc metalloenzyme reaction mechanisms. The aim will be to show how different computational methods provide complementary insight into the biochemistry of these enzymes. We start by noting that many Zn-containing enzymes have been suggested to follow a mechanism similar to that shown in Figure 2, where the role of Zn is to activate the nucleophilic water molecule towards deprotonation (step 1), and to stabilize the negative charge that develops in the tetrahedral intermediate formed upon nucleophilic attack (step 2). In step 3, the tetrahedral intermediate is cleaved. Finally, step (4) is a ligand exchange to return to the initial water-bound complex.

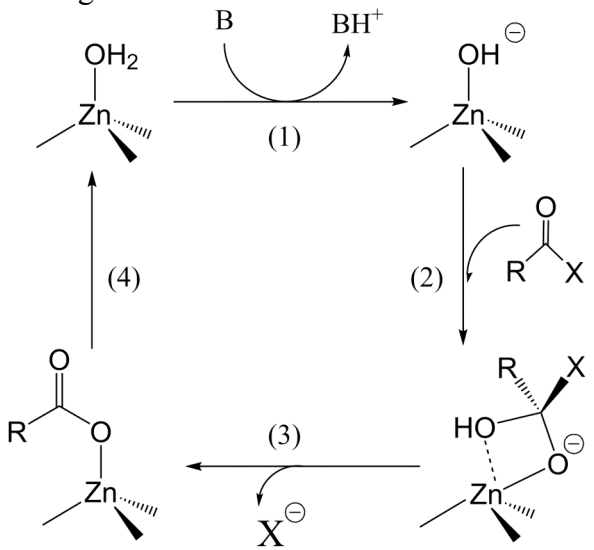

Figure 2 Typical catalytic cycle for Zn-based enzymes.

The capability of the Zn ion to decrease the pK_a of a bound water molecule without decreasing the nucleophilicity of the resulting hydroxide too strongly plays an important role in this overall mechanism, and was the object of a relatively early study[23] using QM calculations. Several model systems with different ligands bound to Zn were considered. In the basic set of models, a Zn ion was bonded to a water molecule or a hydroxide, as well as either three, four or five NH_3 groups (representing His sidechains) to form either tetrahedral, square pyramidal, or octahedral complexes. Other models were generated by replacing one of the NH_3 ligands with either a CH_3S^- group (representing a Cys sidechain), or by an HCO_2^- group (Asp or Glu). Reaction energies for proton transfer from the water molecule were calculated using Hartree-Fock theory. Despite the relatively simple models, and the low level of theory used, some useful trends emerged.

The proton dissociation energy was found to increase steadily with increased Zn coordination. This may explain the preferred tetrahedral coordination in catalytic Zn enzymes, as higher coordination would lead to a less favourable energy

for step 1 in Figure 2. The authors also calculated partial atomic charges, and the energy of the highest energy orbital with oxygen lone pair character, as proxies for the O-nucleophilic character of the molecule. The Zn-bound hydroxide ion was unsurprisingly predicted to be a better nucleophile than the Zn-bound water molecule. The same descriptors led to the prediction that the O-nucleophilicity is not influenced by the coordination mode of Zn. Hence the coordination mode determines the pK_a of the water molecule (step 1 in Figure 2), but has little influence on the nucleophilic character of the ion (step 2 in Figure 2). The authors also used their calculations to account in qualitative terms for the experimental trend in pK_a values of the Zn-coordinated water molecule in the environment of different enzymes, such as carbonic anhydrase, carboxypeptidase and alcohol dehydrogenase.

Turning now to mechanistic studies, we will highlight a few studies focusing on the following five enzymes, whose active sites are presented in Figure 3: carbonic anhydrase, carboxypeptidase A, thermolysin, class B β-lactamases and farnesyltransferase.

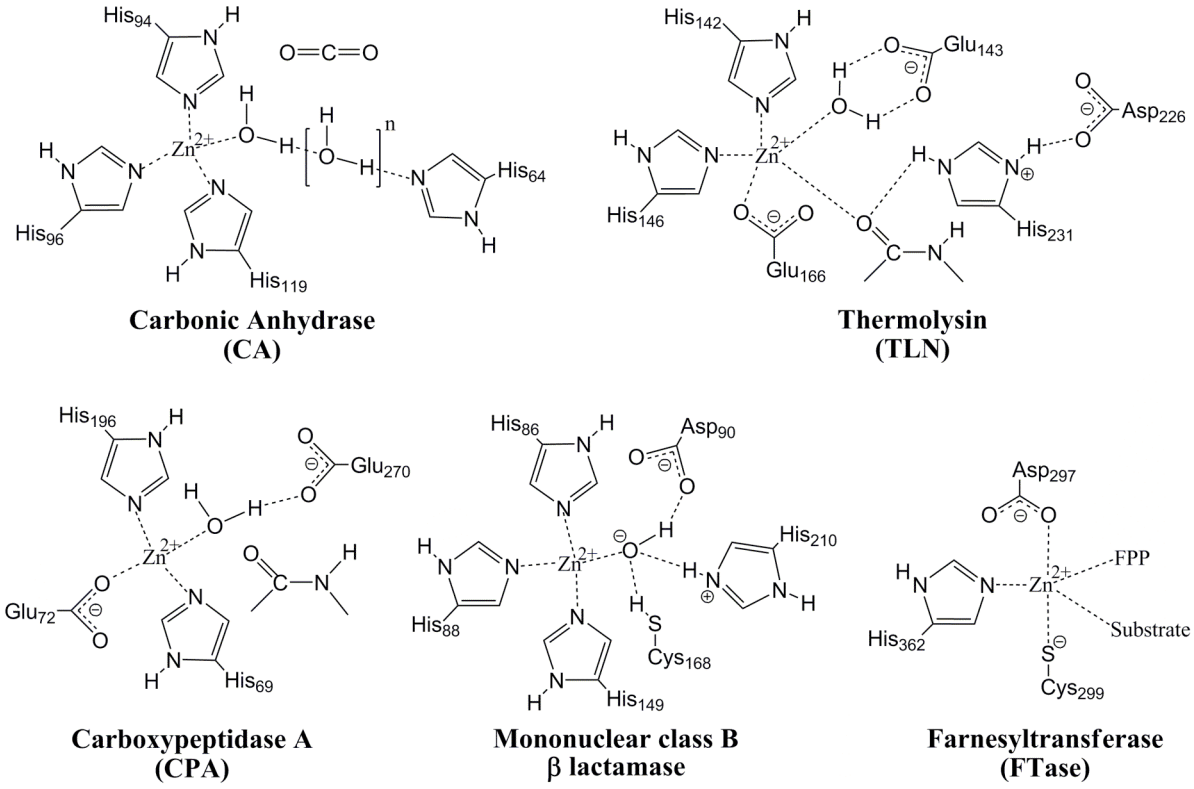

Figure 3 Active site of the five enzymes reviewed in this work.

Carbonic Anhydrase:

Human carbonic anhydrase II (CA) is probably the most heavily studied zinc-containing enzyme, at least in computational terms. A very broad variety of techniques have been applied to CA, over a number of years, and this metalloenzyme thereby provides insight into the development of computational methods in bioinorganic chemistry in the last few decades.

CA is a very efficient enzyme that catalyses the hydration of CO_2 to form HCO_3^-. In the active site, the zinc ion is bound to His94, His96 and His119 and a water molecule (see Figure 3), and the neighboring His64 is also known to play a

catalytically important role. The reaction mechanism of this enzyme closely follows the cycle presented in Figure 2 (with a slight modification for step 3). Many computational studies have addressed the nature of the proton transfer step from the Zn-coordinated water molecule to His64, which was found experimentally to be the rate-limiting step at high buffer concentrations.[24] In typical X-ray structures of the enzyme, the distance between Zn and His64 is around 7.5 Å or greater.[25] This distance is too large for direct proton transfer, so it was proposed that the proton is shuttled from the water molecule to His64 over a "bridge" formed by several water molecules. This suggestion raises two important questions: i) how many water molecules form this bridge, and ii) is the proton transfer concerted or stepwise?

One early computational study of CA was carried out in the group of Lipscomb.[26] In the late 1980s, the available computational resources were rather limited, so only the active site of the system could be considered, using the PRDDO semi-empirical method. A simplified model of the key amino acid sidechains was used (for instance, with NH_3 to represent His sidechains). The calculations led to a predicted catalytic cycle with 5 different steps. These are i) binding of CO_2 near the Zn^{2+}, ii) nucleophilic attack of the hydroxide ion to form HCO_3^-, iii) internal proton transfer in HCO_3^-, iv) binding of a new nucleophilic H_2O to Zn^{2+} and v) proton transfer from Zn-coordinated H_2O to His64 (and then to solvent). Based on these calculations, it was proposed that the proton shuttle from the nucleophilic water molecule to His64 occurs through two bridging water molecules.

The nature of the water bridge has also been studied using MD simulations,[27] of three different protonation states: i) a model of the reactant, with both the Zn-coordinated water molecule and His64 (with a single N atom protonated) treated as neutral entities, ii) a model of the proton-transferred species, with Zn-bound hydroxide ion and a doubly protonated His64 sidechain, and iii) a model for the species in which the proton has been transferred from His64 to bulk solvent. During the ca. 1 ns simulations, the number of water molecules forming the bridge between the nucleophilic water molecule and His64 was monitored, and the results suggested that up to six water molecules may be involved, although for the initial and final stationary points of the proton transfer, four and three bridging water molecules, respectively, were the most common configurations. The water molecules were found to be very mobile, which may explain why experiments do not readily provide information on the length of the bridge. The three simulations also provide insight into how the flexibility of different regions of the protein depend on the protonation state. The three main regions considered were the active site, which is hydrophilic, and two hydrophobic regions above and below it. All three regions were mostly rather rigid, except for the lower hydrophobic region in the case where His64 is protonated. This region is close to His64 and is solvent accessible. The authors suggested that rigidity was generally high in order to maintain the catalytic conformation, but that the greater flexibility of the hydrophobic region upon His64 protonation helps favor proton transfer to the bulk solvent.

Another MD study[28] studied the orientation of His64 during proton transfer. The X-ray structure revealed that there are two possible orientations of this amino acid:[29] the "in" orientation, in which His64 is closer to Zn (ca. 7.5 Å), and the "out" conformation where it is 3 Å further away. MD simulations showed that before the proton is transferred from the Zn-coordinated water molecule to His64, this amino acid prefers to be in the "in" conformation, in order to facilitate the proton transfer, but once the proton has been transferred to His64, this residue changes to the "out" conformation in order to transfer the proton to the solvent.

The use of a QM method to describe the active site allows the proton transfer event itself to be described, as well as the dynamics of the different protonation states. QM/MM PMF simulations using a semi-empirical QM method showed that the calculated free energy barrier for proton transfer from the nucleophilic water molecule to His64 does not depend on the initial orientation of His64.[30] Two different simulations were performed, departing from the "in" and "out" orientations of His64, and both yielded a similar free energy barrier, in the range of 12-13 kcal/mol. In both cases the reaction is found to be concerted: His64 starts to abstract a proton from the closest water molecule, to form a hydroxide ion, and simultaneously, the water molecules in the bridge relay a proton from the Zn-bound water to this incipient hydroxide, leaving a nucleophilic hydroxide ion. This reaction pathway is denoted "proton hole" as there is a tendency for a slightly reduced number of protons to be found on the bridge during the reaction. However, the accuracy of the calculations did not enable the authors to rule out the alternative mechanism, known as Grotthus mechanism, in which proton transfer is initiated at the Zn-coordinated water molecule, which becomes partially deprotonated, yielding an H_3O^+ ion in the bridge, which shuttles the proton to His64.

Carboxypeptidase A:

Carboxypeptidase A (CPA) is a zinc-containing metalloprotease found in the digestive system, which catalyses the scission of bulky C-terminal amino acids from peptide substrates. This enzyme has been extensively studied both experimentally and computationally as a prototype of zinc-containing proteolytic enzymes. The zinc ion is coordinated by the His69 and His196 imidazole groups, by Glu72, and a water molecule (see Figure 3). The substrate is embedded into the active site by strong interactions between its carboxylate terminus and the sidechains of Tyr248, Asn144 and Arg145. Other key amino acids are Glu270, which is found in several X-ray structures to accept a hydrogen bond from the Zn-bound water molecule, and Arg127, which donates a hydrogen bond to the oxygen of the scissile peptide bond of the substrate.[31]

Historically, two different mechanisms have been proposed: the "water promoted pathway"[31] is similar to the generic mechanism of Fig. 2, with the Zn-coordinated water molecule acting as nucleophile. One of its protons is transferred to Glu270 and the resulting hydroxide ion attacks the C atom of the substrate peptide bond to form a gem-diol intermediate. The second part of the reaction involves protonation of the peptide NH group by Glu270, and bond cleavage to yield product. In the alternative denoted as the "anhydride pathway",[32] Glu270 itself acts as a nucleophile to form an acylated carboxylic anhydride intermediate, after peptide bond cleavage.

In an early MD study, the dynamics of a complex between CPA and a typical substrate, the Val-Leu-Phe-Phe tetrapeptide, were explored in order to give insight into which mechanism is more likely.[33] Three relatively short (ca. 100 ps) simulations were run, starting from different structures. In the first, the substrate is located similarly to a substrate analogue found in a crystal structure. The water molecule bound to the zinc ion interacts with Glu270, while the oxygen of the substrate scissile peptide bond interacts with Arg127. In the second simulation, a similar starting structure was used, but with the substrate further away from the active site. Finally, in the last case, the oxygen atom of the peptide bond was bonded to zinc,

replacing the Zn-bound water molecule of the crystal structure. The first two simulations evolved to a very similar equilibrium position, with the same key interactions as in the starting structures, showing that the final conformation of the system is not highly dependent on the initial structure. However, the third simulation led to a different final structure, again very similar to its starting position. The fact that relatively stable simulations can be run in geometries suited for both mechanisms suggests that in structural terms, both processes might be possible.

The energy barriers for the two pathways have been computed for a model system using semiempirical methods.[34] The QM model treated included 120 atoms, i.e. the Zn ion, the substrate, and truncated models of the side chains of the most important surrounding residues. In each mechanism, the rate-limiting step was found to be associated with protonation of the peptide NH group, but the anhydride pathway was found to have a much higher barrier to reaction, 50 kcal/mol above the Michaelis complex, with the barrier for the water-promoted mechanism lying at only 25 kcal/mol. As the authors pointed out, due to the limitations in the model and method used, these calculated energies are of qualitative value only, but the large difference in the two barrier heights suggests that anhydride mechanism can be ruled out.

We have recently studied the water-promoted mechanism using a much more accurate DFT method, in a QM/MM adiabatic mapping procedure.[35] Starting from a Michaelis complex structure with the water and the substrate bound to Zn, a reaction pathway was found with a barrier of ca. 15 kcal/mol, consistent with the experimentally observed reactivity. Proton transfer from water to Glu270 occurs prior to, but in the same step as, nucleophilic attack, and in fact, proton transfer from Glu270 to the substrate peptide NH group also occurs in the same step. Arg127 is found to interact strongly with the negatively charged substrate carbonyl oxygen in the tetrahedral intermediate. The calculated barrier is not sensitive to variations in the DFT method used for the QM region. This study confirms that the water-promoted mechanism should be preferred, and gives some insight into the interactions that stabilize the tetrahedral intermediate.

Thermolysin:

Thermolysin (TLN) is a thermostable enzyme produced by *Bacillus thermoproteolyticus*. Despite its bacterial origin, it resembles many human enzymes, and is an active focus of research as a model zinc-containing enzyme. The active site (see Figure 3) and the reaction catalyzed by the enzyme are similar to those of CPA. The Zn ion is bonded to His142, His146, Glu166, and a water molecule, with Glu143 and His231 also known to have a catalytically important role.[36] TLN catalyses hydrolysis of the N-terminal peptide bond of peptides with a hydrophobic N-terminus. A mechanism related to the general pattern presented in Figure 2 was proposed by Matthews,[37] involving deprotonation of the Zn-coordinated water molecule by Glu143, followed by nucleophilic attack to form a tetrahedral intermediate, proton shuffling, and C-N bond cleavage. An alternative mechanism was proposed[38] in which His231 is the general base instead of Glu143.

The entire catalytic cycle of TLN has been studied using a QM/MM hybrid method, with a semiempirical QM approach.[39] The reaction pathway characterized followed the mechanism proposed by Matthews. The reaction is predicted to take place in a step-wise mode, with calculations finding successive TSs for proton transfer from the nucleophilic water molecule to Glu143, nucleophilic attack on the substrate carbonyl group, protonation of the peptide NH group, and cleavage of the

peptide bond. The rate-limiting step, predicted to be cleavage of the C-N bond, has a barrier of 41.6 kcal/mol relative to the Michaelis complex, much higher than the experimental value. This discrepancy is not unexpected given the limitations of semi-empirical methods. The main benefit of such calculations is to provide insight into the structure of the different stationary points, including the TSs. It is also possible to analyze the contribution of each amino acid to the energy of the different species. One important finding of this type is that His-231 contributes significantly to the stabilisation of the tetrahedral intermediate, providing an alternative explanation for the catalytic role of this residue without involving it directly in proton transfer from the nucleophilic water molecule.

A complementary study used DFT calculations on a cluster model comprising the zinc and key neighbouring groups to explore the mechanism and reactivity for this enzyme.[40] The nature of the reaction pathway as predicted in this study is rather similar to that obtained in the QM/MM study[39] However, the proton transfers from the water to the Glu143 then to the peptide NH group are found to occur simultaneously to the nucleophilic attack, rather than in a separate step. Several models of different size and several methods were compared. In all cases, His residues were modelled as imidazoles, aspartate and glutamates were modelled as formates or acetates (except for Glu143 which was described as a larger butyrate group). The largest model included 68 atoms, namely the three protein side chains bound to Zn, as well as Glu143, His231, Asp226, the nucleophilic water molecule and the substrate (treated as N-methylacetamide). The authors tested the restraining effect of the protein matrix, by freezing an atom of each residue to its position in the crystal structure. Compared to full geometry optimization, this approach was found to lead to rather small structural and energetic effects. The rate-limiting step is found to be the nucleophilic attack, with a barrier of 15.2 kcal/mol, in good agreement with experimental values.[41] In this study too, His231 is found to be important in stabilizing the tetrahedral intermediate through hydrogen bonding, with omission of this residue from the model leading to an increased energy barrier by as much as 8.3 kcal/mol.

Another study of TLN was recently performed by Blumberger et al.,[42] using DFT-based QM/MM methods to run MD simulations and generate a PMF. The large computational expense of the QM/MM simulations restricted the calculations to a few ps for each sampling window. As in the QM cluster model, it was found that approach of the nucleophilic water to the substrate carbonyl group led to concerted deprotonation by Glu143. The computed free energy barrier was 14.8 kcal/mol, similar to that in the QM study. The PMF study suggests however that a range of rather different conformations of the TS and the resulting tetrahedral intermediate are sampled, a situation that cannot easily be detected in a cluster model calculation.

β-lactamases:

These enzymes are produced by some bacteria as a defence mechanism against β-lactam antibiotics, leading to hydrolysis of the lactam ring. There are four classes (A - D) of β-lactamase, of which class B corresponds to Zn-enzymes, which are therefore the sole focus here. These enzymes contain either one or two zinc ions; in the former case, Zn^{2+} is bound to three histidines and a water molecule (see Figure 3). Mutagenesis experiments demonstrate a catalytic role for nearby Asp and His residues. A cysteine residue located in the active site is also important for the catalytic activity of mononuclear β-lactamases. Analysis of the pH-rate profiles for hydrolysis of several substrates suggests that the zinc-bound water molecule is

deprotonated at neutral pH. Several computational studies have attempted to clarify the reaction mechanism in both the mononuclear and binuclear forms.

A QM/MM PMF study of the mechanism of the hydrolysis of cefotaxime by the monozinc β-lactamase from *Bacillus cereus* was carried out using DFT.[43] The active site involves a Zn ion bound to three histidines and a hydroxide ion, which interacts through hydrogen bonds with Asp120, Cys221 and an additional water molecule. The first step is nucleophilic attack of the hydroxide ion on the lactam ring carbonyl group. The calculations suggest that this enzyme involves an unusual mechanistic feature, whereby the second water molecule coordinates to zinc in the same step as nucleophilic attack, stabilising the forming tetrahedral intermediate. The newly coordinated water also facilitates the proton transfers needed before ring opening of the lactam. Finally, after product release, this water molecule becomes the zinc-bound hydroxide ion needed for the next cycle. The computed free energy barriers for nucleophilic attack and ring opening are 18.5 and 21.0 kcal/mol. The statistical error associated with the PMF calculations is of the same order of magnitude as the difference between these barriers, so it was not possible to conclude which step was rate-limiting. The calculations provide a rationale for the residues Asp120, Cys221 and His264 that are known to play a catalytic role but do not directly intervene in the mechanism, as these groups all engage in hydrogen bonding to the hydroxide and the incoming water molecule.

Farnesyltransferase:

Farnesyltransferase (FTase) is involved in receptor tyrosine kinase (RTK) signal transduction pathways, and its malfunction is often associated with uncontrolled cell growth that may lead to cancer formation, so this metalloenzyme has been the target of many anticancer research projects. The reaction catalyzed by this enzyme, thioether formation from a cysteine and a diphosphate (Figure 4), is very different from those described previously. The mechanism is believed to involve four steps: i) binding of the FPP substrate and the protein, ii) activation of the resulting ternary complex, iii) the chemical reaction step, and iv) rate-limiting product release.

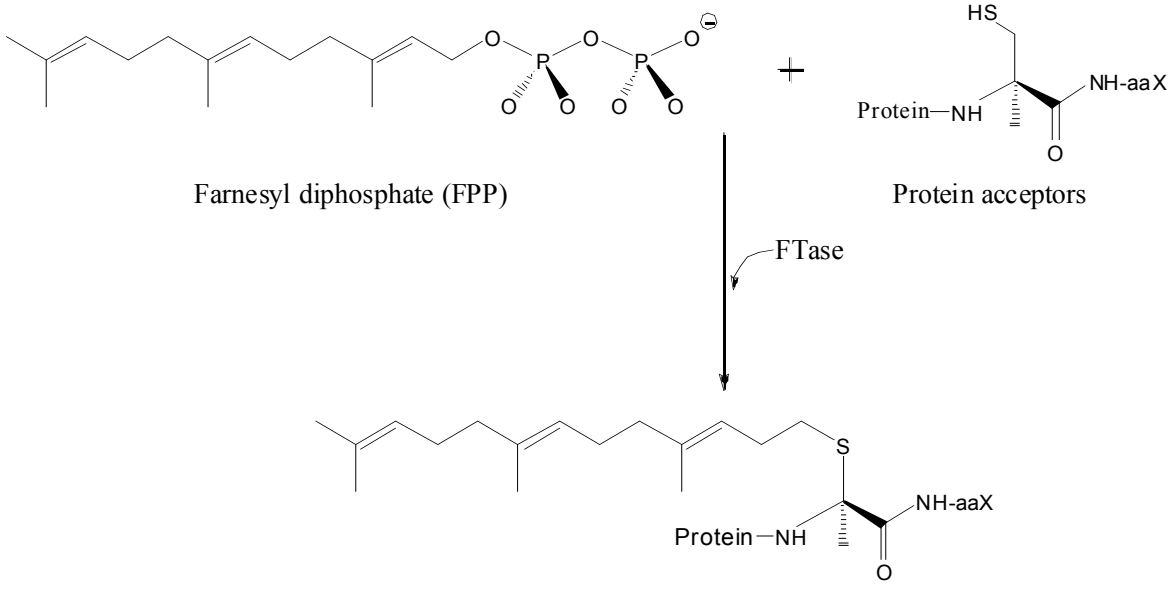

Figure 4 Reaction catalysed by Farnesyltransferase (FTase).

In crystal structures, the distance between FPP and the protein substrate is large, (ca. 7 Å) so this conformation is considered "non-reactive". A conformational change was suggested in which bond torsion occurs in the FPP moiety before reaction occurs, but this hypothetical intermediate has not been identified experimentally. Also, Tyr300 and Lys164 affect the reaction rate, but their precise role is not known. Two MD studies[44, 45] were aimed at understanding the structure and dynamics of the ternary complex. In the initial study,[44] a 5.8 ns MD simulation was run in which the FPP was treated as fully deprotonated (charge of −3), and highlighted the importance of hydrogen bonding between FPP and residues such as Lys164, Arg291, and Lys294. No role for Tyr300 was found. A pH-dependent kinetic study suggested that FPP may be monoprotonated,[46] so the authors repeated their study in this protonation state. The new 8 ns simulation maintained many of the existing hydrogen bond interactions, but also showed a stable hydrogen bond between FPP and Tyr300, that was absent in the previous study. In the same paper,[45] the authors computed a PMF for the approach of the FPP carbon atom to the cysteine side chain of the protein substrate. In the non-reactive ternary complex, this distance is close to 7 Å, but in the case of the monoprotonated FPP, there is a local free energy minimum with a smaller distance of 5 Å, lying only 0.4 kcal/mol above the non-reactive species, and separated from it by a very small barrier. No such minimum is found in the case of the trianionic FPP, suggesting (a) reaction does indeed involve a protonated substrate and (b) conformational change of the FPP leading to closer approach to the cysteine is involved in the reaction mechanism.

3. CONCLUSIONS

Zinc enzymes have been the object of many different computational studies. MD simulations using MM force fields can provide valuable insight into the structure and dynamics of key species such as the Michaelis complex. Such simulations can probe factors such as conformational changes required for reaction. QM and QM/MM calculations can model bond making and breaking: they provide insight into TS structure and energies, and compare different putative mechanisms. Simulations will be increasingly important in studies of zinc enzymes and their mechanisms.

REFERENCES:

1. S. J. Lippard and J. M. Berg, *Hydrolytic enzymes: Carboxypeptidase A and Thermolysin: structural studies*, Mill Valley, California, 1994.
2. D. W. Christianson, *Adv. Protein Chem.*, 1991, **42**, 281.
3. J. J. R. Fraústo da Silva and R. J. P. Williams, in *The biological chemistry of the elements : the inorganic chemistry of life. 2nd ed.*, Oxford University Press., 2001.
4. W. N. Lipscomb and N. Strater, *Chem. Rev.*, 1996, **96**, 2375.
5. J. E. Coleman, *Ann. Rev. Biochem.*, 1992, **61**, 897.
6. B. L. Vallee and D. S. Auld, *Acc. Chem. Res.*, 1993, **26**, 543.
7. A. Warshel, P. K. Sharma, M. Kato, Y. Xiang, H. B. Liu and M. H. M. Olsson, *Chem. Rev.*, 2006, **106**, 3210.
8. T. Dudev and C. Lim, *Chem. Rev.*, 2003, **102**, 773.
9. F. Jensen, *Introduction to Computational Chemistry, 2nd Edition*, Wiley.
10. J. Gao, S. Ma, D. T. Major, N. Kwangho, J. Pu and D. G. Truhlar, *Chem. Rev.*, 2006, **106**, 3188.
11. M. W. van der Kamp and A. J. Mulholland, *Nat. Prod. Rep.*, 2008, **25**, 1001.
12. A. Vedani and D. W. Huhta, *J. Am. Chem. Soc.*, 1990, **112**, 4759.
13. R. H. Stote and M. Karplus, *Proteins, Struc. Func. Gen.*, 1995, **23**, 12.
14. Y. P. Pang, K. Xu, J. El Yazal and F. G. Prendergast, *Prot. Sci.*, 2000, **9**, 1857.
15. J. Aqvist and A. Warshel, *J. Mol. Biol.*, 1992, **224**, 7.
16. M. Dal Peraro, K. Spiegel, G. Lamoureux, M. De Vivo, W. F. DeGrado and M. L. Klein, *J. Struct. Biol.*, 2007, **157**, 444.
17. N. Gresh, J. P. Piquemal and M. Krauss, *J. Comput. Chem.*, 2005, **26**, 1113.
18. D. V. Sakharov and C. Lim, *J. Am. Chem. Soc.*, 2005, **127**, 4921.
19. A. Warshel and M. Levitt, *J. Mol. Biol.*, 1976, **103**, 227.
20. J. Straub, in *Computational Biochemistry and Biophysics*, eds. O. M. Becker, A. D. MacKerell Jr., B. Roux and M. Watanabe, Marcel Dekker, Inc., New York, 2001, p. 199.
21. G. M. Torrie and J. P. Valleau, *J. Comput. Phys.*, 1977, **23**, 187.
22. S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman and J. M. Rosenberg, *J. Comput. Chem.*, 1992, **13**, 1011.
23. I. Bertini, C. Luchinat, M. Rosi, A. Sgamellotti and F. Tarantelli, *Inorg. Chem.*, 1990, **29**, 1460.
24. D. N. Silverman and S. H. Vincent, *CRC Critical Rev. in Biochem.*, 1983, **14**, 207.
25. A. E. Ericksson, A. T. Jones and A. Liljas, *Proteins: Structure, Function, and Bioinformatics*, 1988, **4**, 274.
26. J. Y. Liang and W. N. Lipscomb, *Int. J. Quant. Chem.*, 1989, **36**, 299.
27. S. Toba, G. Colombo and K. M. Merz, *J. Am. Chem. Soc.*, 1999, **121**, 2290.
28. C. M. Maupin and G. A. Voth, *Biochemistry*, 2007, **46**, 2938.
29. K. Hakansson, M. Carlsson, L. A. Svensson and A. Liljas, *J. Mol. Biol.*, 1992, **227**, 1192.
30. D. Riccardi, P. Konig, H. Guo and Q. Cui, *Biochemistry*, 2008, **47**, 2369.
31. D. W. Christianson and W. N. Lipscomb, *Acc. Chem. Res.*, 1989, **22**, 62.
32. J. Suh, J. K. Park and B. K. Hwang, *J. Am. Chem. Soc.*, 1992, **114**, 5141.
33. L. Banci, I. Bertini and G. La Penna, *Proteins*, 1994, **18**, 186.

34. A. Kilshtain-Vardi, G. Shoham and A. Goldblum, *Molec. Phys.*, 2003, **101**, 2715.
35. M. W. Y. Szeto, J. I. Mujika, J. Zurek, A. J. Mulholland and J. N. Harvey, *J. Mol. Struct. Theochem*, 2008, in press.
36. S. Toma, S. Campagnoli, E. Degregoriis, R. Gianna, I. Margarit, M. Zamai and G. Grandi, *Protein Engineering*, 1989, **2**, 359.
37. B. W. Matthews, *Acc. Chem. Res.*, 1988, **21**, 333.
38. W. L. Mock and M. Aksamawati, *Biochem. J.*, 1994, **302**, 57.
39. S. Antonczak, G. Monard, M. R. Lopez and J. L. Rivail, *J. Mol. Model.*, 2000, **6**, 527.
40. V. Pelmenschikov, M. R. A. Blomberg and P. E. Siegbahn, *J. Biol. Inorg. Chem.*, 2002, **7**, 284.
41. K. Morihara and H. Tsuzuky, *Eur. J. Biochem.*, 1970, **15**, 374.
42. J. Blumberger, G. Lamoureux and M. L. Klein, *J. Chem. Theory Comput.*, 2007, **3**, 1837.
43. M. Dal Peraro, L. I. Llarrull, U. Rothlisberger, A. J. Vila and P. Carloni, *J. Am. Chem. Soc.*, 2004, **126**, 12661.
44. G. Cui, B. Wang and K. M. Merz, *Biochemistry*, 2005, **44**, 16513.
45. G. Cui and K. M. Merz, *Biochemistry*, 2007, **46**, 12375.
46. M. J. Saderholm, K. E. Hightower and C. A. Fierke, *Biochemistry*, 2000, **39**, 12398.